RE-425J

STRATEGY SYNTHESIS IN AERIAL DOGFIGHT GAME MODELS

April 1972

RESEARCH DEPARTMENT

GRUMMAN AEROSPACE CORPORATION
BETHPAGE NEW YORK

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Optimal Control Theory Game Theory Mathematical Modeling Feedback Controls						

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Grumman Aerospace Corporation	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

Strategy Synthesis in Aerial Dogfight Game Models

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Research Report

5. AUTHOR(S) (First name, middle initial, last name)

Michael Falco
Victor Cohen

6. REPORT DATE April 1972	7a. TOTAL NO. OF PAGES 37	7b. NO. OF REFS 7

8a. CONTRACT OR GRANT NO. None	9a. ORIGINATOR'S REPORT NUMBER(S) RE-425J
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	None

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY N/A

13. ABSTRACT

The main problem of interest in this report is the "role-definition problem" arising in one-on-one dogfight game models. The computational approach is aimed at providing a decomposition of the space of game initial conditions into sets of unilateral capture capability for each of the players, and at outlining the draw and sacrifice sets in accordance with the players' individual preferences for game outcomes. The procedure develops the feedback policy (in terms of the observable data) that attains the above decomposition. Two highly simplified one-on-one games are considered. The first game model is a discrete time-state alternating move game (perfect information) on a horizontal grid reminiscent of the Isaacs examples. The second model is a continuous time-regional feedback game (imperfect information) in the horizontal plane. The strategy synthesis is effected by a "reinforcement learning" procedure in both game models. Computational results are given in some detail for the first game, while preliminary results are presented for the second game model.

Security Classification

Grumman Research Department Report RE-425J

STRATEGY SYNTHESIS IN AERIAL DOGFIGHT GAME MODELS[†]

by

M. Falco

and

V. Cohen

System Sciences

April 1972

[†] Presented at the Air to Air Combat Analysis and Simulation Symposium, Kirtland Air Force Base, New Mexico, 29 February-2 March 1972. To be published in the Proceedings.

Approved by: Charles E. Mack, Jr
Charles E. Mack, Jr.
Director of Research

STRATEGY SYNTHESIS IN AERIAL DOGFIGHT GAME MODELS

Michael Falco and Victor Cohen

Research Department
Grumman Aerospace Corporation
Bethpage, New York 11714

ABSTRACT

The main problem of interest in this paper is the "role-definition problem" arising in one-on-one dogfight game models. The computational approach is aimed at providing a decomposition of the space of game initial conditions into sets of unilateral capture capability for each of the players, and at outlining the draw and sacrifice sets in accordance with the players' individual preferences for game outcomes. The procedure develops the feedback policy (in terms of the observable data) that attains the above decomposition. Two highly simplified one-on-one games are considered. The first game model is a discrete time-state alternating move game (perfect information) on a horizontal grid reminiscent of the Isaacs examples. The second model is a continuous time-regional feedback game (imperfect information) in the horizontal plane. The strategy synthesis is effected by a "reinforcement learning" procedure in both game models. Computational results are given in some detail for the first game, while preliminary results are presented for the second game model.

INTRODUCTION

One of the more difficult areas for applications oriented workers in the field of modern optimal control theory continues to be the one-on-one aerial dogfight problem. We believe, in this case, these difficulties are due in part to the fact that the one-on-one dogfight problem is perhaps more accurately modeled as a "qualitative" differential game, as contrasted with the "quantitative" game model. Briefly, the qualitative game is such that it contains two or more events dealing with termination of play, for which the players' have some preferential ordering, as contrasted with the quantiative game for which real valued payoff functions defined on the trajectory and/or terminal data can be unequivocally assumed as goals for each player. The Isaacs "homicidal chauffeur game" and "game of two cars" (Ref. 1) are pursuit games of the latter type. In these, the roles of pursuer and evader are clear at the outset, and players seek to minimize (and maximize) the capture time, respectively. Dogfight game models do not come equipped a priori with the pursuer and evader roles defined, in fact these role definitions must be determined in the course of obtaining a resolution of these games.

The approach taken here is a small step in the direction of trying to resolve these dogfight game models. By resclution, we mean to decompose the space of game initial conditions into sets of unilateral capture capability for each player and to outline the sacrifice and draw sets in accordance with the players individual preferences for game outcomes, and furthermore to derive the associated strategies (providing the decomposition) as feedback control policies on the collection of observable data. Two highly simplified game models are considered in the text. The first is a discrete time-state game with an alternating move structure. The second model is a continuous time-state game model employing "regional" feedback policies. In the case of the first model, "perfect information" regarding the "state" at each player's control decision has been assumed. A resolution of that game model for specified dynamics, control capabilities, weapons envelopes, and player preferences is obtained by two procedures. The first procedure is similar to that employed by Isaacs (dynamic programming) in the homicidal chauffeur game, but with some modification to observe the stipulated preference descriptions of the dogfight instead of the min max capture time criteria of the chauffeur game. The second procedure employs a "reinforcement rule" algorithm used in conjunction with the simulation of game plays. The second procedure offers the conceptual facility for immediate extension to the more complex problem presented by the second model. The second model, as constructed, does not have a predetermined move structure (simultaneous or alternating), but instead the control reevaluation points on a time scale are implicitly determined by the traversing of "regional" boundaries in the observables during the course of play. This "imperfect information" model is similar in many respects to the one constructed by Baron et al. (Refs. 2, 3) in their "controllable" Markov chain approach to pursuit-evasion problems. The text will outline a "reinforcement rule" procedure to be applied in these models as originally described in Ref. 4, and present some preliminary computational results for specific model data.

DISCRETE TIME-STATE DOGFIGHT GAME

Game Model: Description of State, Lethal Envelopes

The state relative to Player I is given by the triple (n,m,p). The admissible control choices for any (n,m,p) for Player I are u_1, u_2, u_3 (see Fig. 1); for Player II are v_1, v_2, v_3 (see Fig. 2). We assume the game to have an alternating move structure. The one step transition equations for a move by Player I are

$$\begin{bmatrix} n \\ m \\ p \end{bmatrix}_{K+1} = \begin{bmatrix} n \\ m \\ p \end{bmatrix}_{K} + \begin{bmatrix} -\left(\frac{n+m}{k_1} + 1\right) & -1 & \left(\frac{m}{k_1} - 1\right) \\ n/k_1 & 0 & -\left(\frac{n+m}{k_1}\right) \\ 1/k_1 & 0 & -1/k_1 \end{bmatrix} \begin{bmatrix} \\ u \\ \\ \end{bmatrix}$$

where K denotes the time unit and where if $p = \pm 3$ (see Fig. 3) and if $u = u_3$, set $p = -3$; or if $u = u_1$, set $p = -3$.

The one step transition equations for a move by Player II are

$$\begin{bmatrix} n \\ m \\ p \end{bmatrix}_{K+1} = \begin{bmatrix} n \\ m \\ p \end{bmatrix}_{K} + \begin{bmatrix} f(q-1) & f(q) & f(q+1) \\ f(p-1) & f(p) & f(p+1) \\ -1/k_2 & 0 & 1/k_2 \end{bmatrix} \begin{bmatrix} \\ v \\ \\ \end{bmatrix} .$$

In the above, $q = p + 2$ and

$$\begin{aligned} f(x) &= +1 \quad \text{if} \quad x = +1, +2 \\ &= -1 \quad \text{if} \quad x = -1, -2, +4, +5 \\ &= 0 \quad \text{if} \quad x = 0, \pm 3, +6 \end{aligned}$$

also if $p = \pm 3$ and if $v = v_2$ or v_3, set $p = -3$; and if $v = v_1$, set $p = +3$. The quantities u and v are interpreted as follows:

When

$u = u_1$ $\qquad$ $u = u_2$ $\qquad$ $u = u_3$

then ; then ; then

$$\begin{bmatrix} \\ u \\ \\ \end{bmatrix} = \begin{bmatrix} k_1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} \\ u \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ k_1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} \\ u \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ k_1 \end{bmatrix} .$$

Similarly, when

$v = v_1$ $\qquad$ $v = v_2$ $\qquad$ $v = v_3$

then ; then ; then

$$\begin{bmatrix} \\ v \\ \\ \end{bmatrix} = \begin{bmatrix} k_2 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} \\ v \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ k_2 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} \\ v \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ k_2 \end{bmatrix} .$$

k_1 and k_2 are step sizes in the grid with which the players move and are representative of the velocities with which grid points can be traversed.

Game Outcome Description

In general, for this game only one of four possible outcomes can result from a play of a game beginning from any (n,m,p). The outcomes are:

C_I	capture by Player I
C_{II}	capture by Player II
S	sacrifice (mutual capture)
D	draw

Note: We have assumed that first "passage" to any of the outcomes C_I, C_{II}, or S terminates play.

On the basis of the lethal envelopes illustrated in Figs. 1 and 2, the sets become:

$$A_I = \left\{ n,m \;\middle|\; \begin{array}{c} 0 \leq n+m \leq 2 \\ 0 \leq n \leq 2 \end{array} \right\}$$

$$A_{II} = \left\{ n,m,p = 0 \;\middle|\; \begin{array}{c} -3 \leq n \leq 0 \\ -3 \leq n+m \leq 0 \end{array} \right\}$$

$$= \left\{ n,m,p = 1 \;\middle|\; \begin{array}{c} -3 \leq m \leq 0 \\ -3 \leq m+n \leq 0 \end{array} \right\}$$

$$= \left\{ n,m,p = 2 \;\middle|\; \begin{array}{c} -3 \leq m \leq 0 \\ 0 \leq n \leq 0 \end{array} \right\}$$

$$= \left\{ n,m,p = 3 \text{ (or } -3) \;\middle|\; \begin{array}{c} 0 \leq n \leq 3 \\ 0 \leq n+m \leq 3 \end{array} \right\}$$

$$A_{II} = \left\{ n,m,p = -2 \quad \begin{matrix} 0 \le m \le 3 \\ 0 \le n+m \le 3 \end{matrix} \right\}$$

$$= \left\{ n,m,p = -1 \quad \begin{matrix} -3 \le n \le 0 \\ 0 \le m \le 3 \end{matrix} \right\} .$$

Hence the sets

$$C_I \triangleq A_I \cap \overline{A_{II}}$$

$$C_{II} \triangleq A_{II} \cap \overline{A_I}$$

$$S \triangleq A_I \cap A_{II}$$

$$D \triangleq \overline{A_I \cup A_{II}}$$

dealing with termination can be described in terms of the (n,m,p) coordinates.

Move Structure and Information Pattern

We have postulated an alternating move structure in this discrete game. Therefore, the move structure and information patterns fall into one of the two game structures shown in Fig. 4, where the argument of $x(\cdot)$ and $u(\cdot)$, $v(\cdot)$ is the time unit.

We assume the move structure and information pattern of Game I (e.g., Player I moves first) in subsequent discussion. The game move structure is interpreted as follows: the game begins at $x(0)$ (coordinates n,m,p). Player I has complete information, that is, knowledge of (n,m,p) at the time he makes control decision $u(0)$. The game state is advanced via the transition equations to state $x(1)$, at which point Player II, having data $x(1)$, selects decision $v(1)$, and so on, until a termination occurs. At this point, we require a stopping time parameter, T, from which a draw termination can be decided in a fixed number of stages of play.

Strategies

The strategies for this game are the functions ζ, η, where:

$$\text{for Player I} \quad x(N) \xrightarrow{\zeta} u(N)$$

$$\text{Player II} \quad x(N) \xrightarrow{\eta} v(N) \quad .$$

Hence, ζ is a mapping from all $x(N)$ to an admissible u (likewise for η and v), and the totality of all ζ, (η) the strategy spaces. N is the index of time (or stage) of play. In our algorithm we utilize behavior strategies, and the actual choice of move made at $x(N)$, is then accomplished by sampling from the stipulated distribution.

Outcome Preferences

In line with our treatment of dogfight games as qualitative games, there exists a preference for outcomes C_I, C_{II}, S, and D on the part of each of the players. For this example, a typical preference ordering might be given as:

Player I	C_I	preferred to	D, S, C_{II}
	D	preferred to	S, C_{II}
	S	preferred t o	C_{II}
Player II	C_{II}	preferred to	D, S, C_I
	D	preferred to	S, C_I
	S	preferred to	C_I

Computational Approach Using Reinforcement Rule Logic

Model Assumptions Made for Computational Expediency

- Truncation of the game state to a finite collection. The truncation is such that the region shown by the shading in Fig. 5 represents the finite collection of states, while the region exterior to it constitutes termination as a draw outcome. In realistic models, this boundary would be representative of those relative range values at which visual or other contact could not be made. In our model, therefore, we consider that any path, even though it starts in the interior, upon reaching the exterior is terminated as a draw.

- Introduce a fixed termination time that terminates all paths as draw outcomes beyond the fixed time. This time is a parameter of the model and can be varied to examine the solution's dependence on the values of this parameter.

- Strategies are functions of the current state only and not time (or time-to-go) and state.

The Simulation Process

- Data

1) Indexing of the finite state $1, \ldots, \tilde{N}$.

2) Dynamical system: one stage reachable set description given for Players I and II.

3) Classification of outcomes: sets C_I, C_{II}, S, in terms of weapons system descriptors A_I and A_{II}.

4) Termination time specified: T.

5) Probability distributions on control choices initially equally likely for all states for both players.

6) Subjective reinforcement rule weightings assigned to outcomes $\Omega = \{C_I, C_{II}, S, D\}$ in accordance with given orderings; weightings $\mu(\Omega)$ for Player I; $\nu(\Omega)$ for Player II.

- **Obtaining A Run**

1) An initial game state is selected. A random number generator is consulted for determination of control choice. The sampling is done in accordance with the probability distributions currently used by that player for that state. Hence, a pair of state-control sequences are generated.

$x(0), u(0) \quad ; \quad x(2), u(2) \quad ; \quad \ldots$

$x(1), v(1) \quad ; \quad x(3), v(3) \quad ; \quad \ldots$

These data are temporarily stored. An outcome is observed, say C_I; the run is then terminated. Assume the arbitrary weights

$\mu(C_I) = 2.00 \qquad \nu(C_{II}) = 2.00$

$\mu(D) = 1.00 \qquad \nu(D) = 1.00$

$\mu(S) = 0.99 \qquad \nu(S) = 0.99$

$\mu(C_{II}) = 0.5 \qquad \nu(C_I) = 0.50$

have been assigned. (These weights are in accordance with the example ordering given earlier.)

2) The reinforcement process is conducted as follows:

For Player I: Assume state x_i visited u_1 chosen by I at x_i. Hence the distribution at x_i is altered from $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

For Player II: Assume x_i visited, v_2 chosen by II at x_i. Hence the distribution at x_i is altered from $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to (2/5, 1/5, 2/5).

This procedure is repeated for states visited during that run by both players. Note: This is an arbitrary procedure; other possibilities exist, one being to alter the distributions nearer termination more than those nearer the start of that run. This is a point for further investigation and is incorporated in the continuous model. Hence for the procedure described we change the distributions in the following way: Let $n_1(x_i)$, $n_2(x_i)$, $n_3(x_i)$ represent nonnegative entries for Player I associated with state x_i. Initially, $n_1 = n_2 = n_3$; hence

$$\text{Prob}[u(x_i) = u_K] = \frac{n_K}{\sum_{j=1}^{3} n_j} .$$

As we have assumed that C_I was the termination, then the new entries become

$$\mu(C_I)n_1(x_i),\ n_2(x_i),\ n_3(x_i) \quad ,$$

since u_1 was utilized by Player I when x_i was the current state. These quantities are then normalized and used as new data for obtaining the next run of the simulation. (A similar procedure is carried out for Player II.)

At this point in time, our experience with the above model is not sufficient to disclose the most efficient sampling procedure over the game starting conditions nor the most efficient reinforcement rule logic. However, our experience has shown that building from short duration games from starting points close to termination outward to longer duration games from more distant starting points (s_ lar to dynamic programming) is a preferred procedure with the reinforcement rule mentioned.

The Markov Chain Models

As our interest in these problems is to obtain a decomposition of the game starting conditions into sets for which

Player I can capture Player II, II can capture I, and sets of mutual capture, according to the players' respective outcome preferences, the following Markov chain model proves useful:

- The transition operators of Markov Chains are described first. We assume that a sufficient number of runs have been made in the simulation process and that two families of stable distributions representing the strategies for Players I and II over the x_i have been obtained.

  For Player I we can then form P where

$$P = \begin{array}{c|cccccccc|}
 & x_1 & x_2 & x_3 & x_j\cdot\cdot & x_k\cdot\cdot & x_\ell\cdot\cdot & x_{\bar{N}} & x_{\bar{N}+1} \\
\hline
x_1 & 1 & 0 & 0 & & & & & \\
x_2 & 0 & 1 & 0 & & & & & \\
x_3 & 0 & 0 & 1 & 0 & & & & \\
\vdots & & & & & & & & \\
x_i & & 0 & 0 & p_{ij} & p_{ik} & p_{i\ell} & & \\
\vdots & & & & & & & & \\
x_{\bar{N}} & & & & & & & & \\
x_{\bar{N}+1} & & & 0 & 0 & 0 & 0 & & 1 \\
\hline
\end{array}$$

where

$$p_{ij} = \text{Prob}[x(K+1) = x_j \mid x(K) = x_i] \quad .$$

In the above, we have let

$$x_1 = \left\{ x \mid x \in C_I \right\}$$

$$x_2 = \left\{ x \mid x \in C_{II} \right\}$$

$$x_3 = \left\{ x \mid x \in S \right\} .$$

We then require $p_{11} = 1$, $p_{22} = 1$, $p_{33} = 1$ by our first passage assumption. The entries for arbitrary row x_i $(p_{ij}, p_{ik}, p_{i\ell})$ are obtained from two sources: 1) the numerical value p_{ij} from the converged distributions in the strategy table for the corresponding control choice; and 2) the location j,k,ℓ from the one-step reachable set properties of the dynamical system of Player I. The state space truncation to a finite collection N with termination as draw outside this collection is treated by the additional state $x_{\bar{N}+1}$ with the property that

$$p_{\bar{N}+1,\bar{N}+1} = 1.0 \quad .$$

A similar construction is used to obtain an operator Q for Player II analogous to P for Player I.

- Given the operators P and Q, we can now compute the following conditional probability of entrance:

$$\text{Prob}\left\{ x(K) = x_1,\ x(\nu) \neq x_2, x_3 \mid x(0) = x_i \right\}$$

where

$$0 \leq K \leq T$$

$$0 \leq \nu < K \quad ,$$

and where T is the stopping time parameter. Hence, we have the probability that play will first terminate in C_I in T stages or less, given that play began at $x(0) = x_i$. These data are obtained in the first column of the matrix $[PQ]^T$ in game I (Player I moving first) and in $[QP]^T$ in game II (Player II moving first). The second column signifies termination in C_{II}, the third column in S. These probability data serve to provide the decomposition sought.

Computational Results

Figures 6-9 show the decomposition obtained for game I with $k_1 = 2$, $k_2 = 1$, $T = 10$ moves for each player, the dynamics, lethal envelopes, and player preferences all being assumed as outlined in the previous discussion. The plots for $p = -1$, $p = -2$ are not shown as these data are available from their symmetrical counterparts $p = +1$, and $p = +2$, respectively.

Note: One finds that all strategies are pure strategies in the converged results as might have been expected from the alternating move - perfect information structure of the problem. The detailed listing of the associated strategies for both players making up the decomposition is not given, because of space considerations.

Computer Considerations

The above described procedure was programmed for use on an IBM 360/75 computer. The model was composed of 2166 states, (n,m,p) triples, by means of equivalence class reductions in the terminations of type C_I, C_{II}, S; the resulting state was reduced to $\bar{N} = 2046$ (symmetric conditions could have reduced this figure by nearly half). A total of 50,000 runs (plays) were made in arriving at the strategy distributions. This required 20 minutes of computer time. The conditional probability of entrance computations used roughly two minutes of computer time to obtain the above decomposition. Symmetry considerations could have reduced the running times to 12 minutes for the example above.

Storage requirements were as follows for the above problem:

Strategies (probability distribution as floating point) One Stage Reachable Set (integer packing) Simulation Routine with Reinforcement Rule Logic	100,000 bytes (4 bytes per word)

$$\left.\begin{array}{l}\text{Conditional Probability of Entrance}\\ \text{Computations Using Markov Chain Model}\end{array}\right\}\quad \text{120,000 bytes}$$

The computer utilized has a 500,000-byte core capacity.

A Second Computational Procedure

In this section, we briefly describe a procedure similar to that used by Isaacs (Ref. 1) (in solving the discrete chauffeur game) and apply it to the discrete time-state dogfight model. This procedure has special merit in this perfect information - alternating move model in that the decomposition of game initial conditions in accordance with the player preference orderings is accomplished with minimal computational expense.

The procedure is as follows:

1) Given termination data C_{I_0}, C_{II_0}, S_0 (subscript here refers to number of moves by I to termination in C_I, C_{II}, S). Given preference ordering for outcomes for individual players.

2) Initialize array

$$\begin{array}{ll|c|c|c|c|c|c|}
 & & \multicolumn{6}{c}{\text{Control}} \\
 & & u_1 & u_2 & u_3 & v_1 & v_2 & v_3 \\
\hline
 & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\text{State} & x_i & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
 & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 & & & & & & & \\
\end{array}$$

```
3)  Select $x_i \notin C_{I_o} \cup C_{II_o} \cup S_o$

    a)  for $u_1$  if $x_i \xrightarrow{u_1} x \in C_{I_o}$
                set $x_i, u_1 = 1.0$ in array

            if $x_i \xrightarrow{u_1} x \in C_{II_o}$
                set $x_i, u_1 = 0$ in array

            if $x_i \xrightarrow{u_1} x \in S_o$
                set $x_i, u_1 = 0.3$ in array

            if $x_i \xrightarrow{u_1} x \in D$
                set $x_i, u_1 = 0.7$ in array

    b)  Do a) over all $u_j$

    c)  For $x_i$
        (1) if $\exists$ at least one $u_j = 1.0$ in array
            for that row call $x_i \subset C_{I_1}$

        (2) if $\exists$ no $u_j = 1.0$ and at least one
            $u_j = 0.7$ $x_i$ is not labeled
        (3)  if $\exists$ no $u_j = 1.0$, and no $u_j = 0.7$,
             and at least one $u_j = 0.3$ call
             $x_i \subset S_1$
        (4)  if $\exists$ no $u_j = 1.0$, and no $u_j = 0.7$,
             and no $u_j = 0.3$ call $x_i \subset C_{II_1}$

4)  Do step 3) over predetermined range of
    $x_i \notin C_{I_o} \cup C_{II_o} \cup S_o$

5)  Select $x_k \notin C_{I_o} \cup C_{II_o} \cup S_o \cup C_{I_1} \cup C_{II_1} \cup S_1$
    if $x_\ell \in C_{II_o} \cup C_{II_1}$ set $x_k, u_1 = 0$ go to 5 d)
    if $x_\ell \in S_o \cup S_1$ set $x_k, u_1 = 0.3$ go to 5 d)

    a)  For $u_1, v_1$: if $x_k \xrightarrow{u_1} x_\ell \xrightarrow{v_1} x_m$
        (1)  if $x_m \in C_{I_o} \cup C_{I_1}$
             set $x_\ell, v_1 = 0$ in array
        (2)  if $x_m \in C_{II_o} \cup C_{II_1}$
             set $x_\ell, v_1 = 1$ in array
        (3)  if $x_m \in S_o \cup S_1$
             set $x_\ell, v_1 = 0.3$ in array
        (4)  if $x_m \in D$
             set $x_\ell, v_1 = 0.7$ in array

    b)  Do a) over $v_j$
```

```
c) For $x_k$
    (1) if $x_m \in C_{I_o} \cup C_{I_1}$ for all $v_j$
        set $x_k,u_1 = 1.0$
    (2) if $x_m \in C_{II_o} \cup C_{II_1}$ for at least one
        $v_j$ set $x_k,u_1 = 0$
    (3) if $x_m \in D$ for at least one $v_j$ and
        $\notin C_{II_o} \cup C_{II_1}$ for any $v_j$ set
        $x_k,u_1 = 0.7$
    (4) if $x_m \in S$ for at least one $v_i$ and
        $\notin C_{II_o} \cup C_{II_1} \cup D$ for any $v_j$ set
        $x_k,u_1 = 0.3$
d) Do a), b), c) for $u_i$
e) For $x_k$
    (1) if $x_k,u_i = 1.0$ for any entry $u_i$
        call $x_k \subset C_{I_2}$
    (2) if $x_k,u_i = 0$ for all entries $u_i$
        call $x_k \subset C_{II_2}$
    (3) if $x_k,u_i \neq 1.0$ for any $u_i$, and
        $x_k,u_i = 0.7$ for at least one $u_i$
        call $x_k \subset D$
    (4) if $x_k,u_i \neq 1.0$ and $x_k,u_i \neq 0.7$ for
        any $u_i$ and $x_k,u_i = 0.3$ for at least
        one $u_i$ set $x_k \subset S_2$
6) Do step 5) over predetermined range of
   $x_k \notin C_{I_o} \cup C_{II_o} \cup S_o \cup C_{I_1} \cup C_{II_1} \cup S_1$
```

7) The extension to 3 and more stages of play using steps 5) and 6) is straightforward.

Note: A simultaneous move version of the discrete-time state model presented is currently under study in the Grumman Research Department. In this case, a revision of the preference ordering (from that assumed here) has been made to obtain ultimately a game for which a zero-sum payoff property is specified. In this case, one of the players is required to prefer the sacrifice outcome over the draw result. A dynamic programming procedure is being used to conduct the strategy synthesis with the optimal mixed strategies in the single-stage subgames determined by a Brown-Robinson iteration procedure. This procedure was first outlined by Kopp (Ref. 5) in the context of a simpler simultaneous move dogfight game model.

CONTINUOUS-TIME-DISCRETE REGION GAME MODEL IN THE HORIZONTAL PLANE

The Continuous-Time "Regional" State One-On-One Aerial Combat Model in the Horizontal Plane

The model for combat in the horizontal plane is a logical extension of the discrete model and thus permits qualitative comparison. Both vehicles are assumed to have constant velocity.

System Equations

The kinematic equations are similar to those given by Isaacs (Ref. 1) for the game of two cars. The equations are written in terms of a coordinate system centered on Player I (see Fig. 10), and are given as

$$\dot{x} = - \frac{V_I}{R_I} y\phi + V_{II} \sin \theta$$

$$\dot{y} = \frac{V_I}{R_I} x\Phi - V_I + V_{II} \cos \theta$$

$$\dot{\theta} = - \frac{V_I}{R_I} \Phi + \frac{V_{II}}{R_{II}} \psi$$

$$-1 \leq \Phi \quad , \quad \psi \leq 1$$

where

V_I and V_{II} are the speeds of Vehicles I and II, respectively;

Φ and ψ the control variables for I and II, respectively (both bounded); and

R_I and R_{II} are the minimum turn radii of I and II, respectively

with $\rho = \sqrt{x^2 + y^2}$ (Range), ω (Bearing), and θ (relative heading angle between V_I and V_{II}).

Observable Data and Control Variables

Since we are interested in constructing feedback controls, $\Phi(\rho,\omega,\theta)$ and $\psi(\rho,\omega,\theta)$, let us look at a proposed decomposition of the visual sphere (or circle and rays in this two dimensional version). Based on discussions with experience combat pilots, we do not believe that relative range, bearing, or heading can be measured accurately in the dogfight encounter. Thus, the state of one aircraft with respect to the other is imperfectly known. To model this imperfect information, we ascertained in a cursory way what is capable of being known and to what degree of accuracy. These discussions led to the partitioning of the visual sphere (or visual horizontal plane in this two dimensional version) as shown in Fig. 11. This partitioning is made with the assumption that Systems I and II are representative of aircraft in the dogfighting situation. The divisions themselves, such as Region 41 in Fig. 11, is meant to imply that Player I can only discern that Player II is somewhere between 6000 and 12000 feet ahead and somewhere between 0° and 7½° off to his right. In the partitioning shown in Fig. 11, the shaded region denotes the lethal gun envelope of I and the region in which I uses a gunsight for lead-pursuit tracking. We have assumed that a lingering time of 0.5 seconds continuously or 1.0 cumulative seconds in the gun envelope constitutes a "kill;" this is a modification of the instantaneous "kill" property of the discrete game. The second player is assumed to have a similar partitioning of the space.

The partitioned state space in ρ and ω has a third coordinate, θ, which we are assuming again to be imperfect. We assume also that θ is known only to lie within the values specified below for Regions 1-41 and that it is not discernible for $\rho > 12{,}000$ feet wherein a vehicle would appear at best as a black dot on the horizon. Hence, θ is observable within the following:

$$315° < \theta \leq 45° \qquad \theta_1$$

$$45° < \theta \leq 135 \qquad \theta_2$$

$$135 < \theta \leq 225° \qquad \theta_3$$

$$225 < \theta \leq 315° \qquad \theta_4 \quad .$$

A similar breakdown applied to Player II. Hence, in this model we have

$$\begin{array}{r} 41 \\ \underline{\times\ 4} \\ 164 \end{array} + 11 = 175 \text{ regions in the decomposition.}$$

We have limited the admissible controls to be finite in number (i.e., $\phi = \pm 1, 0,$ and similarly $\psi = \pm 1, 0$), hence, the probabilistic feedback law would be represented by the following table of state doubles $X_i(R,\theta)$, where R is the region and θ is the relative heading angle between I and II.

For Player I

State	Prob $\phi = +1$	Prob $\phi = -1$	Prob $\phi = 0$
X_1 $(R_1 = 1,\ \theta = \theta_1)$	$p_{1,+1}$	$p_{1,-1}$	$p_{1,0}$
X_2 $(R_1 = 1,\ \theta = \theta_2)$			
⋮			
$X_{164}(R = 41,\ \theta = \theta_4)$	$p_{164,+1}$	$p_{164,-1}$	$p_{164,0}$
$X_{165}(R = 42,\ \text{all } \theta)$			
⋮			
$X_{175}(R = 52,\ \text{all } \theta)$	$p_{175,+1}$	$p_{175,-1}$	$p_{175,0}$

where $p_{164,-1}$ is the probability of choosing the control -1 when Player I discerns that Player II is in Region 164 with respect to himself.

For Player II, we have a similar table with the states given by the proximity of Player I with respect to Player II.

State	Prob $\psi = +1$	Prob $\psi = -1$	Prob $\psi = 0$
X_1 $(R_1 = 1,\ \theta = \theta_1)$	$p_{1,+1}$	$p_{1,-1}$	$p_{1,0}$
⋮			
$X_{175}(R = 52,\ \text{all } \theta)$	$p_{175,+1}$	$p_{175,-1}$	$p_{175,0}$

The sets of capture C_I, C_{II}, and sacrifice S cannot necessarily be identified in terms of ρ,ω,θ at the outset, even though one may be in the envelope of the other, due to the linger time stipulation.

Simulation Procedure

Assume that a family of games is played with durations $0 < T_1 < T_2 < \ldots < T_n$ (see Fig. 12). Assume that the game begins at initial conditions ξ (say in Region 23 for I, corresponds to 52 for II) and has duration T_1. [We select initial conditions close to termination for I (and II).]

Choices of control are selected from $X(R = 23, \theta = \theta_1)$ for Player I and $X(R = 52, \text{all } \theta)$ for Player II. Say, for argument's sake, that they are $\phi = +1$ and $\psi = +1$, respectively. The differential equations are integrated from ξ using $\phi = +1$ and $\psi = +1$ until Region 23 for I or Region 52 for II is exited. If either occurs (or both), the new region for that player is consulted (say $23 \rightarrow 22$ for I, II continues with 52); hence $X(R = 22, \theta = \theta_1)$ is consulted for the next control decision for I which, say, is $\phi = 0$. We continue in this way until an outcome C_I, C_{II}, S, or T_1 is observed. Meanwhile, the "state-control" pairs have been temporarily stored. For

$$\text{I} \quad 23,\ \phi = +1 \ ;\ 22,\ \phi = 0 \ ;\ \ldots$$

$$\text{II} \quad 52,\ \psi = +1 \ ;\ \ldots$$

As in the discrete model, the reinforcement rule is applied to alter the distributions with respect to the temporarily stored data. We have modified the reinforcement rule to be other than multiplication of the control choice chain during any one run by a constant and then normalizing. We have incorporated a linear weighting that reinforces the control choice chain more strongly after many plays of the game, hopefully avoiding the reinforcement of a basically poor choice of control that may have led to a successful outcome on the part of one player because the second had not yet learned how to play adequately. We repeat this procedure over many ξ in regions close to termination using time parameter T_1; T_2 is then selected, and experiments repeated over ξ in regions not previously covered by experiments using T_1.

Note: In this model we do not have to decide whether the game is of simultaneous or alternating move structure; the sequence of moves in time resolves itself in accordance with the assumed decomposition among the observables and the integration of the kinematic equations. It should also be noted that we have used the y-axis as a reflecting barrier and thereby, by symmetry, have reduced the number of stored states in our feedback representation, and subsequently in our simulation.

Preliminary Computational Results

The results presented for the continuous 2-D model are by no means complete, but these results do indicate that the reinforcement algorithm developed for the discrete game carries over directly to the continuous one.

Region 22 (θ_1) as shown in Fig. 13 (and designated simply as 22 in Fig. 11) is considered representative of a region close to termination. We are seeking to ascertain the control policy probability distributions on the part of both players for encounters that begin therein. We are also seeking the probability of the various possible outcomes, C_I, C_{II}, S, and D. We fix the converged control policies for Region 22 (θ_1) and, knowing the probabilistic outcomes for play entering that region, go on to consider Region 22 (θ_2). We start play in the latter region and terminate play if we enter Region 22 (θ_1), which has been previously decided, or terminate by the occurrance of one of the possible outcomes prior to entering Region 22 (θ_1). We reinforce accordingly, and begin new encounters until the control choice probability distribution becomes invariant for Region 22 (θ_2).

The particular parameters that were chosen in this 2-D continuous model were V_I = 1000 ft/sec, V_{II} = 500 ft/sec, R_I = 3000 ft, and R_{II} = 2500 ft. Investigation of the time that any one play from a given initial condition lasts, before a draw is considered the outcome, resulted in a time of 100 seconds. At a relative velocity between the two players of 500 ft/sec this time is sufficient for the faster player to catch the slower if the slower is near the edge of the visual threshold, as shown in Fig. 11, and headed in the same direction.

The primary question to which we addressed ourselves was: What is the most favorable probability distribution on the choice of control decisions for Player I when he finds Player II in Region 22 (θ_1)? Note that even though II is always in Region 22 (θ_1) with respect to I, I is not necessarily in the same region with respect to Player II at the beginning of play. We utilize 80 particular sets of initial conditions; these are specified as all combinations of $\rho = 3600$ ft, 4200 ft, 4800 ft, 5400 ft; $\omega = 1.5°$, 3.0°, 4.5°, 6.0°, and $\theta = -44°$, -22.5°, 0°, 22.5°, 44°, all of which fall into Region 22 (θ_1) of II/I (Player II with respect to Player I).

We begin by dividing the unit interval equally into 80 parts, with each part corresponding to one of the 80 ρ,ω,θ triples (initial conditions). Starting from a uniform distribution on the control policies of Player I for Region 22 (θ_1), we select an initial condition randomly, run a game, observe the outcome, make the reinforcement accordingly, and choose another initial condition; then a game is run, etc., etc. This resulted in a single distribution for the region which was $p_{LT} = 1.0$, $p_{SA} = 0$, and $p_{RT} = 0$, where p_{LT} = probability of making a Left Turn, p_{SA} = probability of going Straight Ahead, and p_{RT} = probability of making a Right Turn. The results of running 1000 random initial conditions chosen from the 80 allowable yielded $p_{C_I} = 0.885$, $p_{C_{II}} = 0.030$, $p_S = 0$, and $p_D = 0.085$. Many of the draw outcomes and captures by Player II occurred during the first few hundred games. If we look at games 500 through 1000, the $p_{C_I} = 0.940$, $p_{C_{II}} = 0.020$, $p_S = 0$, $p_D = 0.040$, which looks very good for Player I. One might conjecture that a left turn when the opponent is ahead and slightly to the right is not the best policy; but after tracing a few of the plays through, one sees that Player I turns left as a delaying maneuver and then right (II/I is in Region 23 (θ_1) or in Region 23 (θ_2) as he turns right) since he has a closing velocity of 500 ft/sec. If he had gone straight, Player II would have turned left and could have held I in the weapons envelope as he passed II. If he turned right, Player II could have made a much sharper right and obtained a draw. Using a different random number generator for selecting the initial conditions and the control choices led to $p_{C_I} = 0.940$, $p_{C_{II}} = 0.009$, $p_S = 0$ and $p_D = 0.051$, but the control policy for Player I converged to $p_{LT} = 0$, $p_{SA} = 1.0$, $p_{RT} = 0$ which tends to indicate that making a left turn or going straight ahead on the part of Player I are equally good policies and result in a high probability of capture. Player II's control policy choice for the initial condition at the end of 1000 games was virtually a uniform distribution in both cases, indicating that all choices of control on his part were equally bad due to his being beaten so many times. Other regions converged during these 1000 runs such as Regions 23 (θ_1) and 23 (θ_2) which converged to $p_{LT} = 0$, $p_{SA} = 0$, and $p_{RT} = 1.0$.

The procedure at this point is to take the resulting distribution tables for each player and start play in an adjacent region such as 22 (θ_3) [since Region 22 (θ_2) had converged to $p_{LT} = 1.0$, $p_{SA} = 0$, $p_{RT} = 0$ in the prior run] and allow the distributions to change. Note that those regions for which the probability distribution on the controls has gone to 1, 0, 0 can never be altered by this algorithm. We can also terminate play when one of those regions, such as Region 22 (θ_1), from which we have already simulated play, is entered since we already know the outcome which began in that region.

CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

It is clear at this point that the general solvability of realistic one-on-one dogfight game models is far from being an accomplished fact. In reality, it is not clear at present that any single computational approach today would have the requisite efficiency and capacity to handle the variety of detailed game models, in which veteran combat pilots might place an ultimate faith. Despite this, there is a great deal of information of a general nature that can be gained with these simple models. For example, obtaining the decompositions of the game initial conditions in a systematic way can lead to parametric studies involving: 1) vehicle parameters; 2) weapon systems parameters; 3) observable data changes; and 4) player preference ordering changes, etc. In this way, the improved capability due to a vehicle-weapons system's change can be directly measured by the "volume" increase of space of initial conditions for which that system has unilateral capture capability; or as might be the case, with improvements in the observable data, improvements in the capture probabilities as well. The associated strategies for attaining these decompositions would also be obtained when making these studies. An additional use for such simple models and their resolution may be to provide the more complex and extremely detailed digital simulation efforts, with the approximate location of the boundaries making up the initial condition decomposition and the associated strategies. The computational method presented here was utilized in a simplified form and although the results sought were obtained, the algorithm as applied in these game models is computationally inefficient. Efforts are underway to devise better sampling procedures and more sophisticated reinforcement learning rules in these models.

REFERENCES

1. Isaacs, R., Differential Games, Wiley, New York, 1965.

2. Baron, S. et al., "A New Approach to Aerial Combat Games," NASA CR-1626, October 1970.

3. Baron, S. et al., A Study of the Markov Game Approach to Tactical Maneuvering Problems, Report No. 2179, Bolt, Beranek, and Newman, Inc., October 1971.

4. Falco, M., An Algorithm for Strategy Synthesis in Aerial Dogfight Game Models, Grumman Aerospace Corporation Research Proposal RP-396, February 1971 (Proprietary).

5. Kopp, R. E., The Numerical Solution of Discrete Dynamic Combat Games, RM-523J, November 1971. Presented at the 4th Inter. Fed. for Info. Processing Coll. on Optimization Techniques, Santa Monica, California, October 17-22, 1971.

6. Starr, A. W., "Nonzero-Sum Differential Games: Concepts and Models," Harvard University Division of Engineering and Applied Physics Technical Report 590, June 1969.

An additional reference and excellent bibliographic source on the topic of differential games and its applications is:

Ciletti, M. D. and Starr, A. W., Differential Games: A Critical View, Differential Games: Theory and Applications (Symposium Proceedings of the AACC Theory Committee), ASME, June 1970.

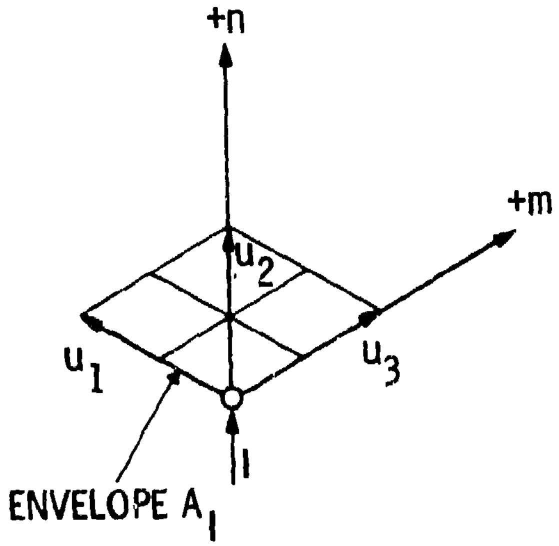

Fig. 1 Lethal Envelope Player I

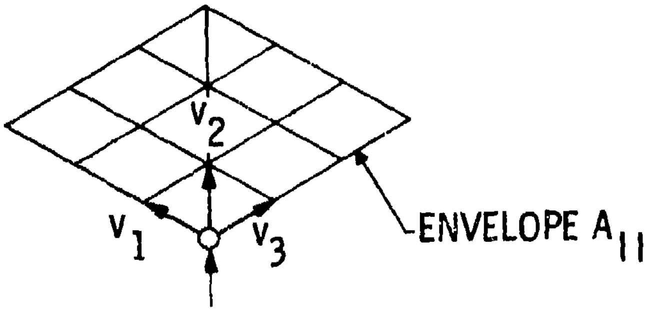

Fig. 2 Lethal Envelope Player II

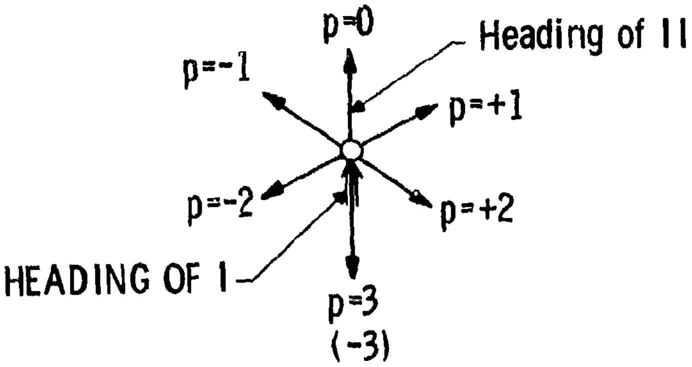

Fig. 3 Relative Heading

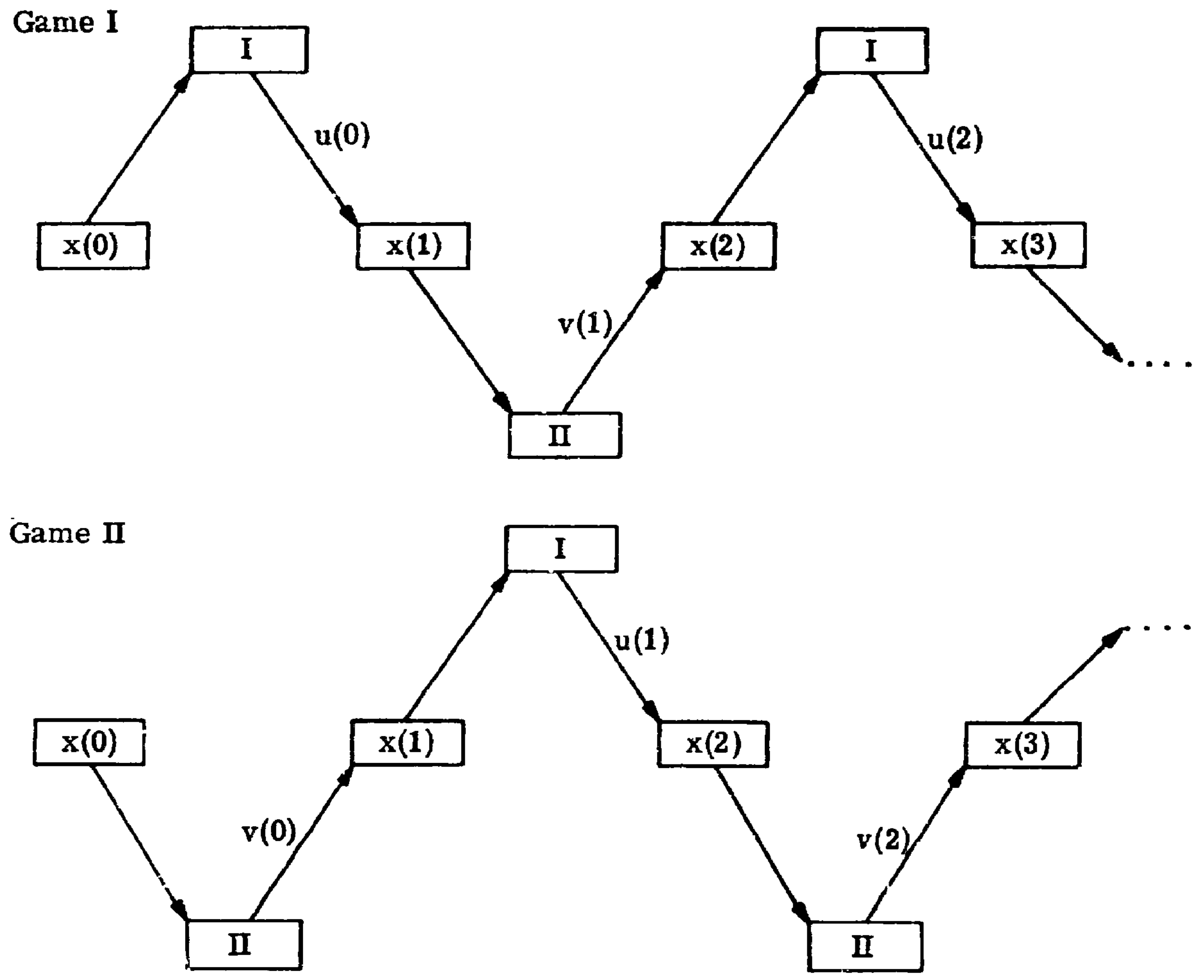

Fig. 4 Move Structure and Information Pattern

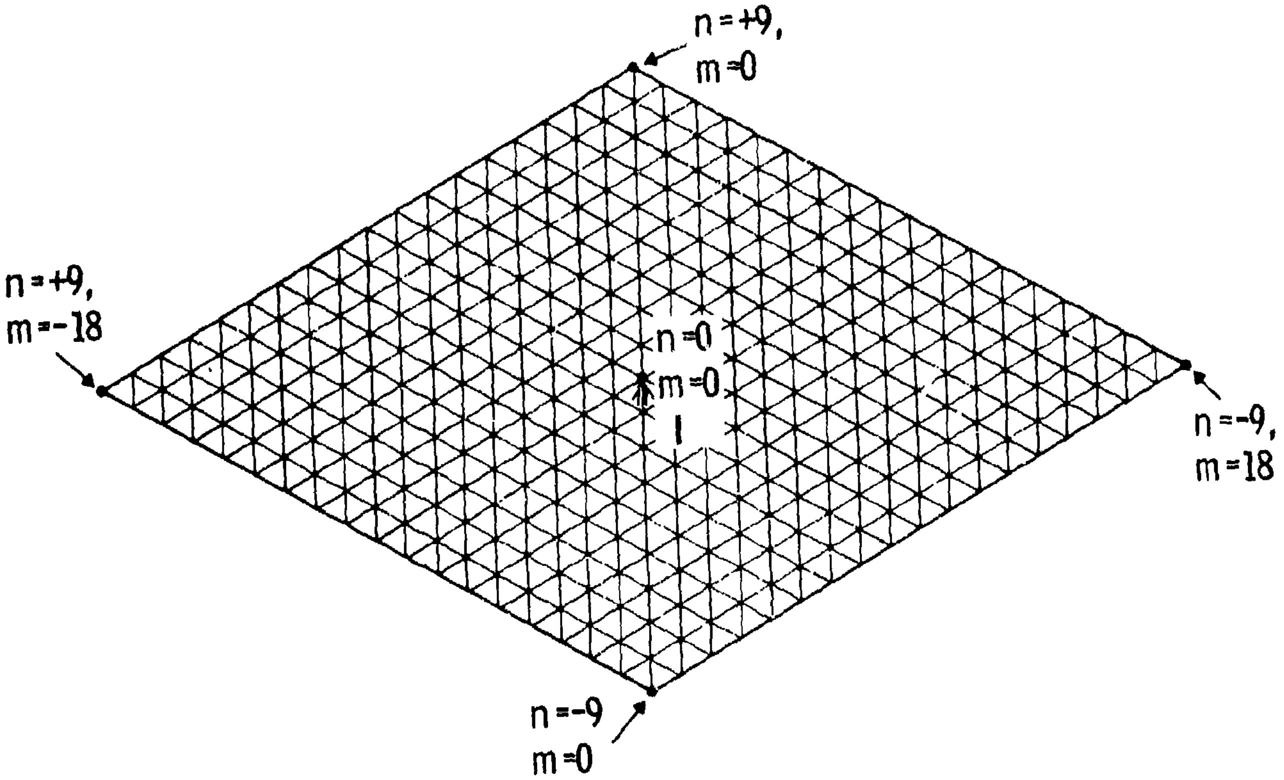

Fig. 5 Truncation of Region of Play

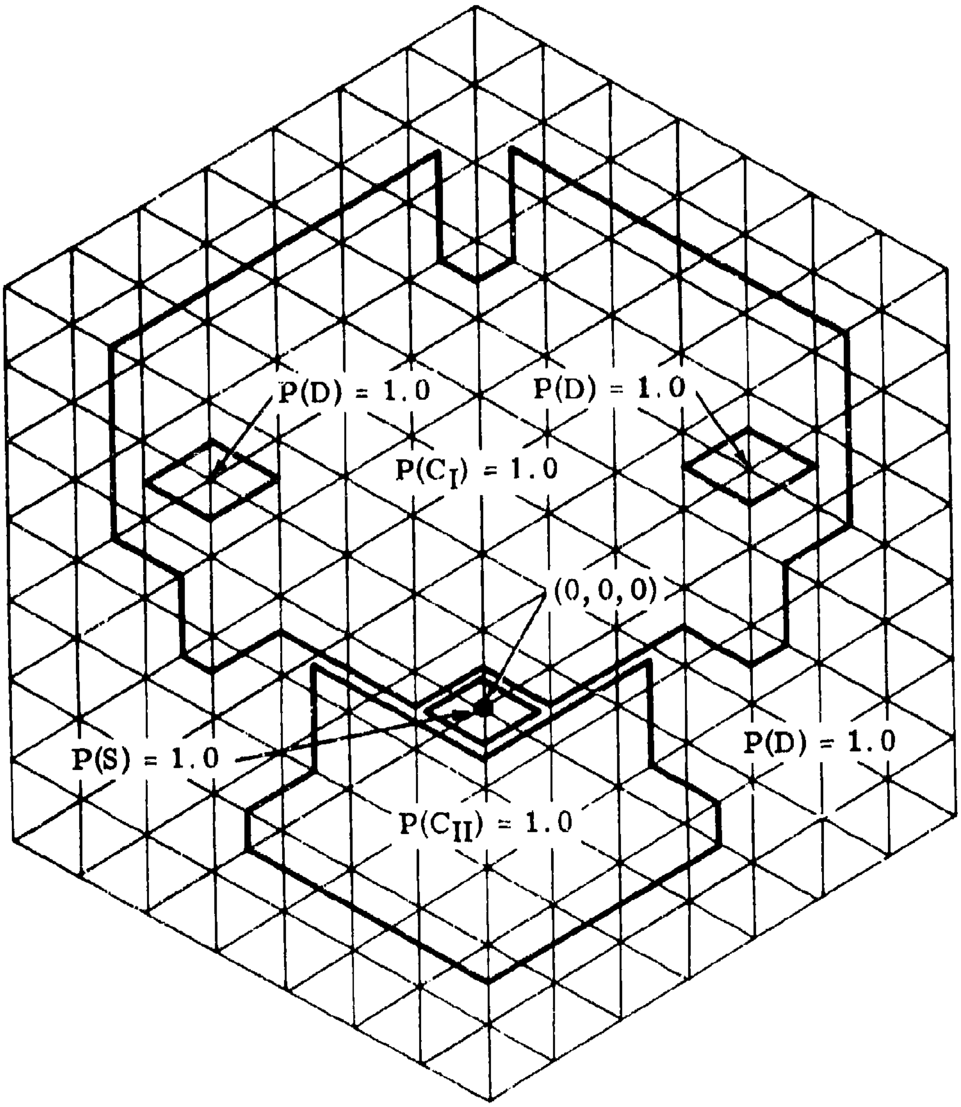

Figure 6. Decomposition of Starting Conditions for All N, M, with P=0

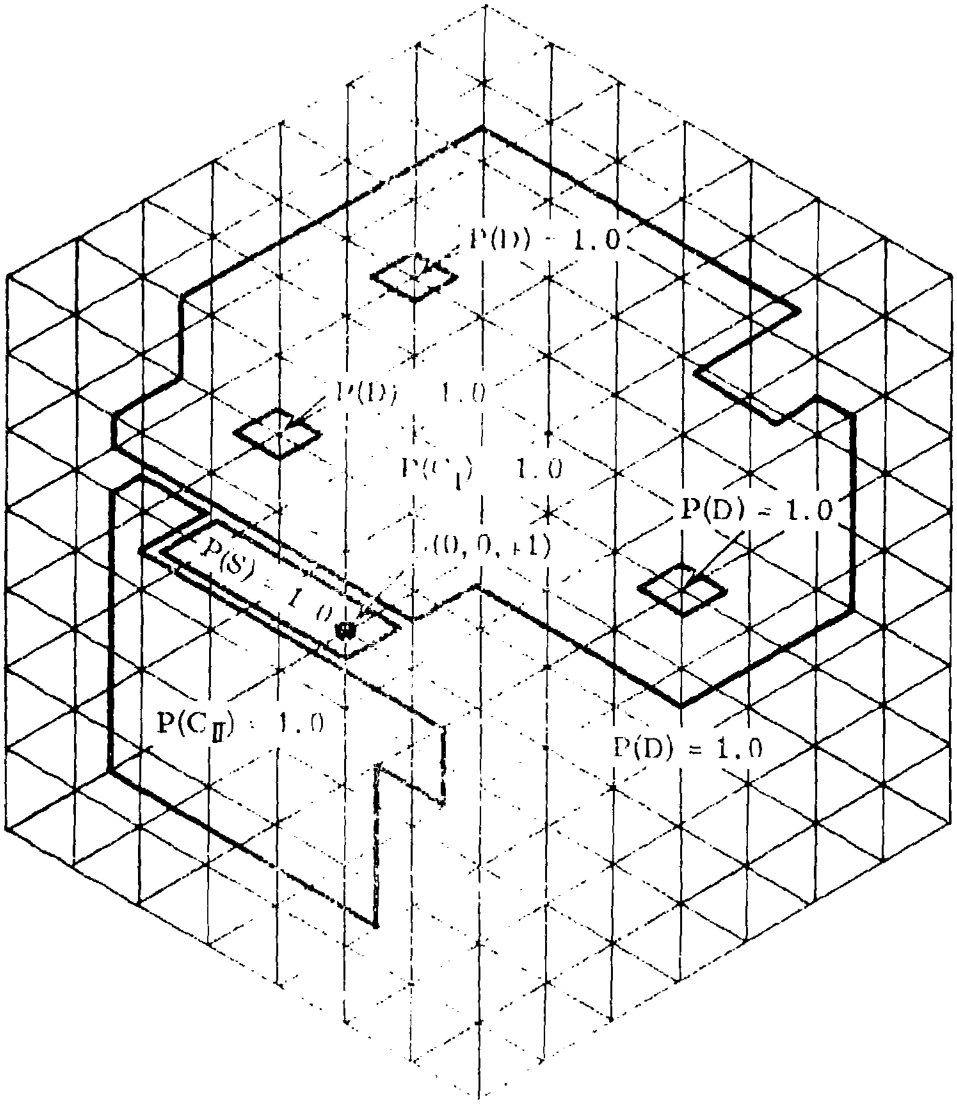

Figure 7. Decomposition of Starting Conditions for All N, M, with P = +1

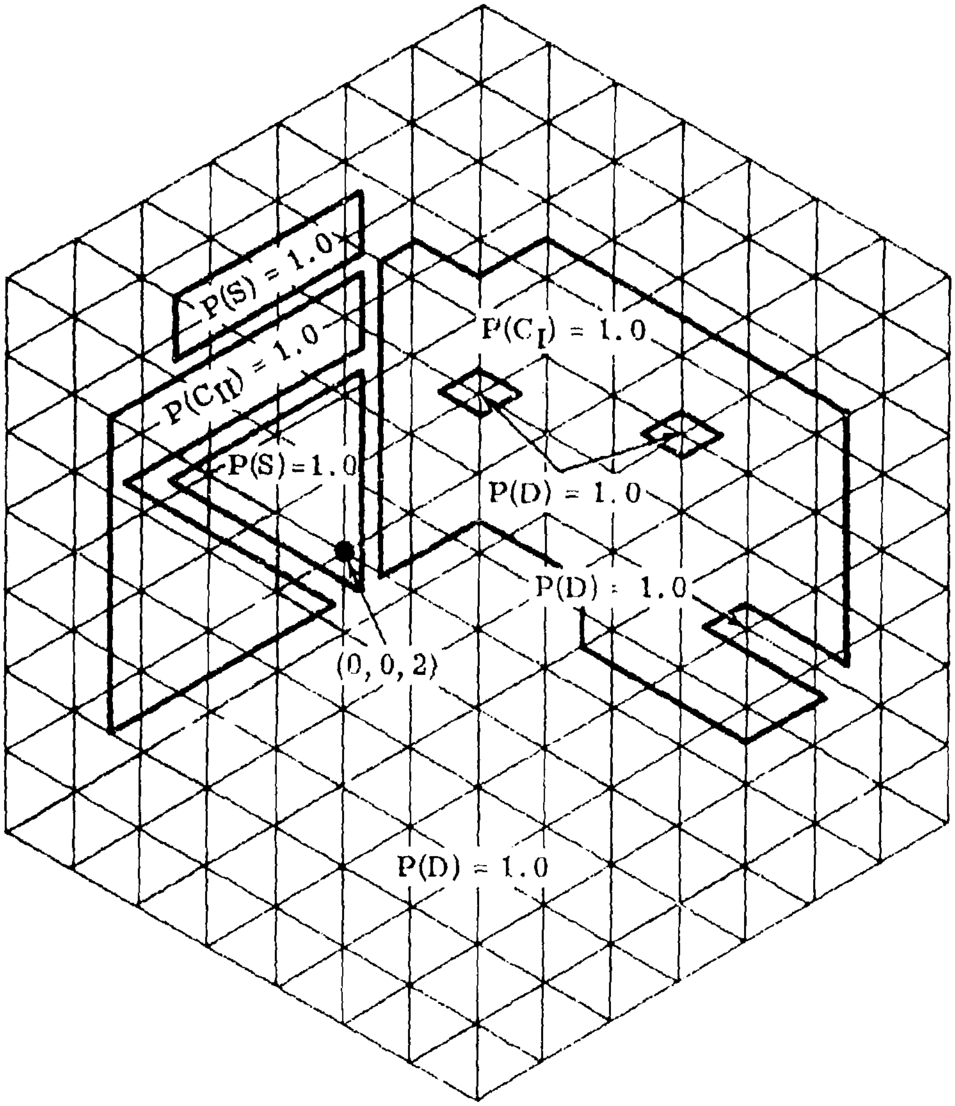

Figure 8. Decomposition of Starting Conditions for All N, M, with P = +2

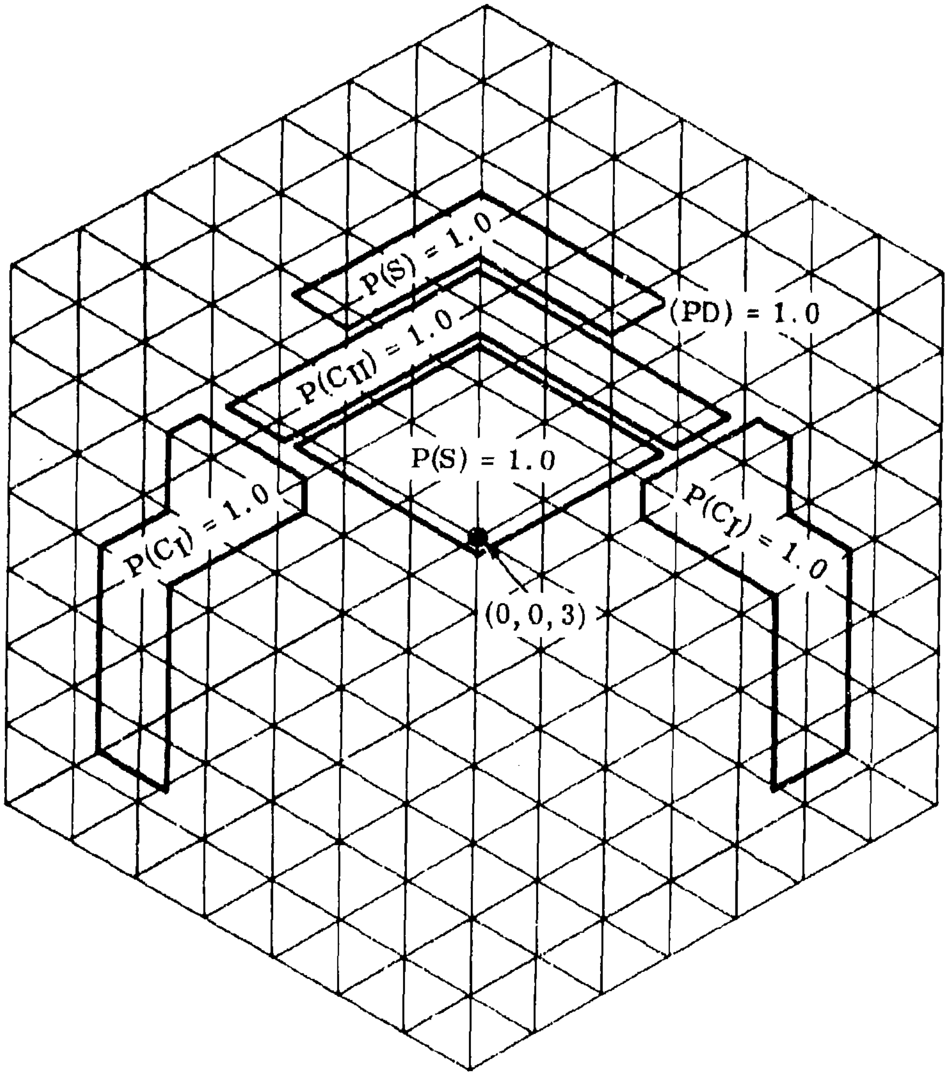

Figure 9. Decomposition of Starting Conditions for All N, M, with P = 3

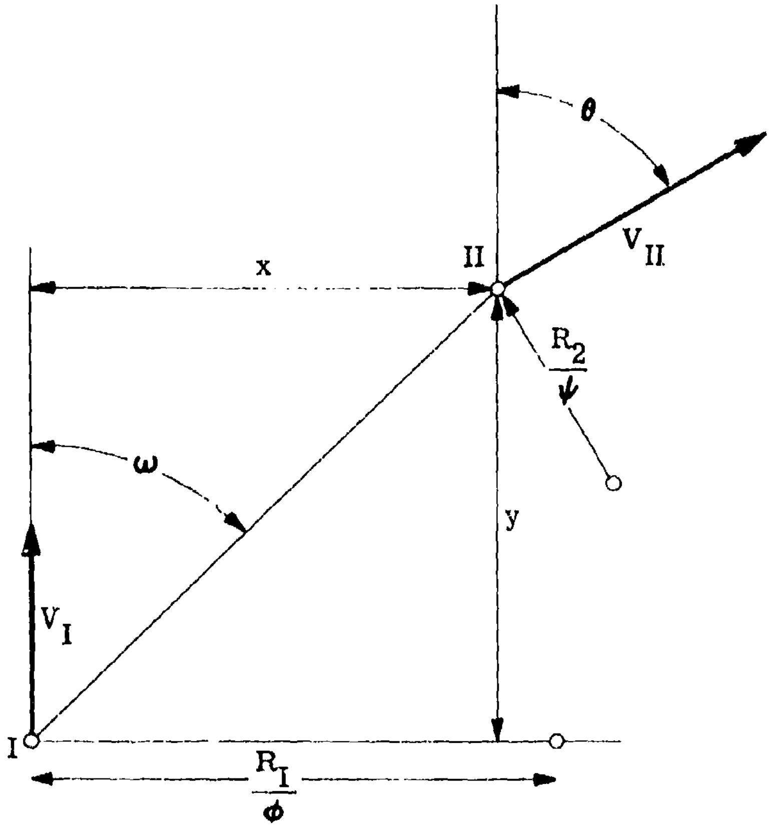

Fig. 10 Coordinates for Game in Horizontal Plane

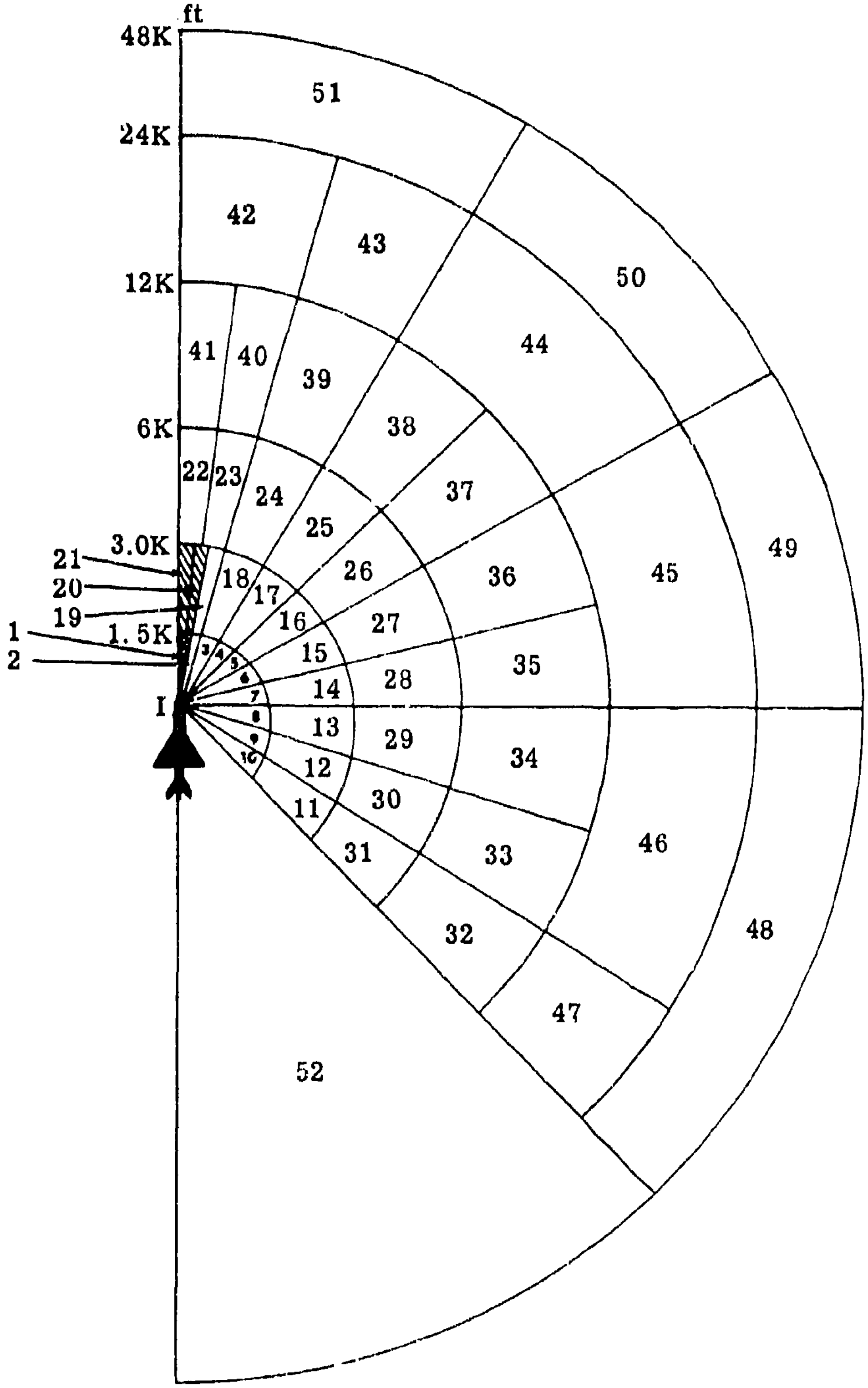

Fig. 11 Regions in the Horizontal Plane

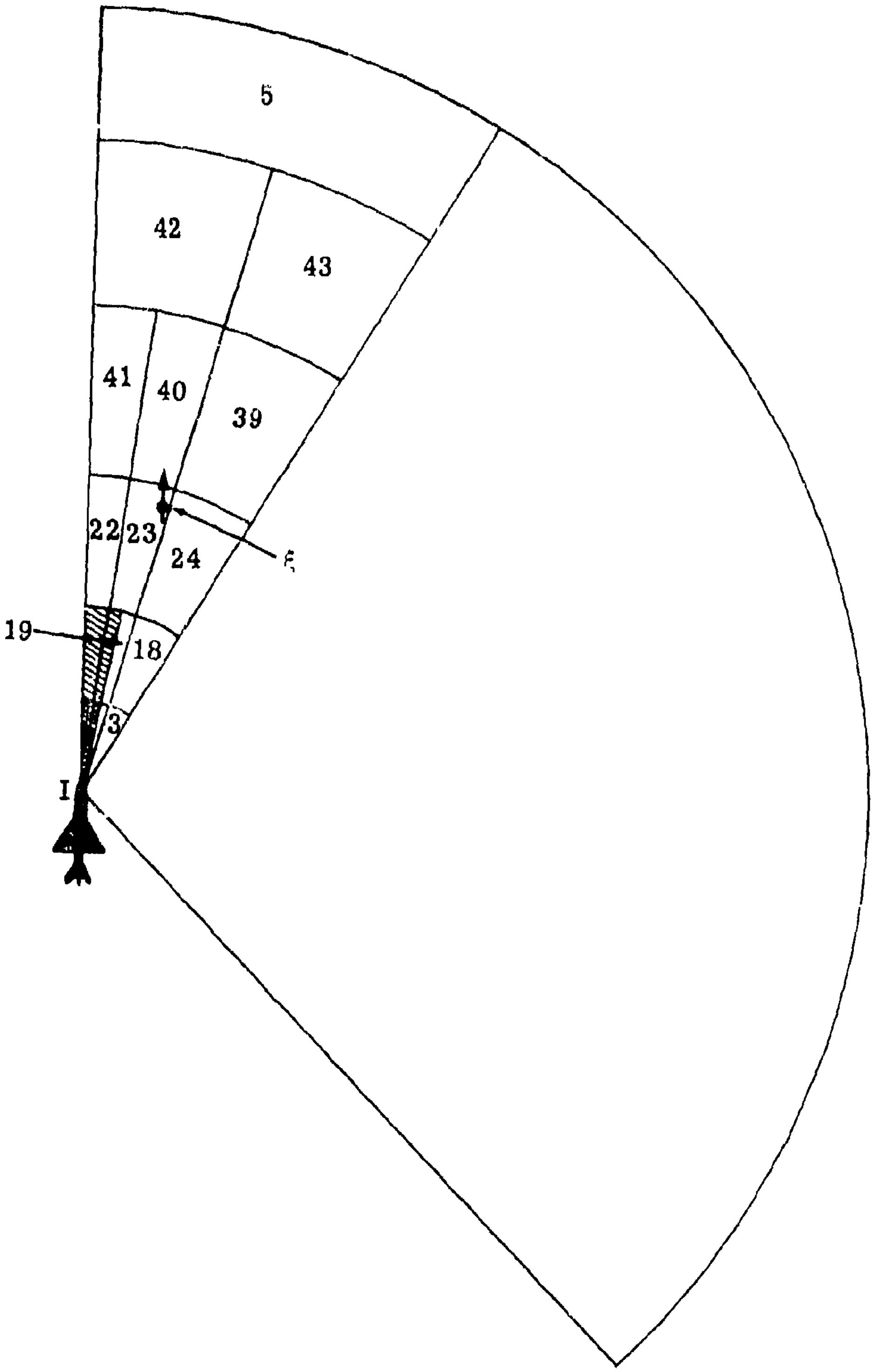

Fig. 12 Game Play in the Horizontal Plane

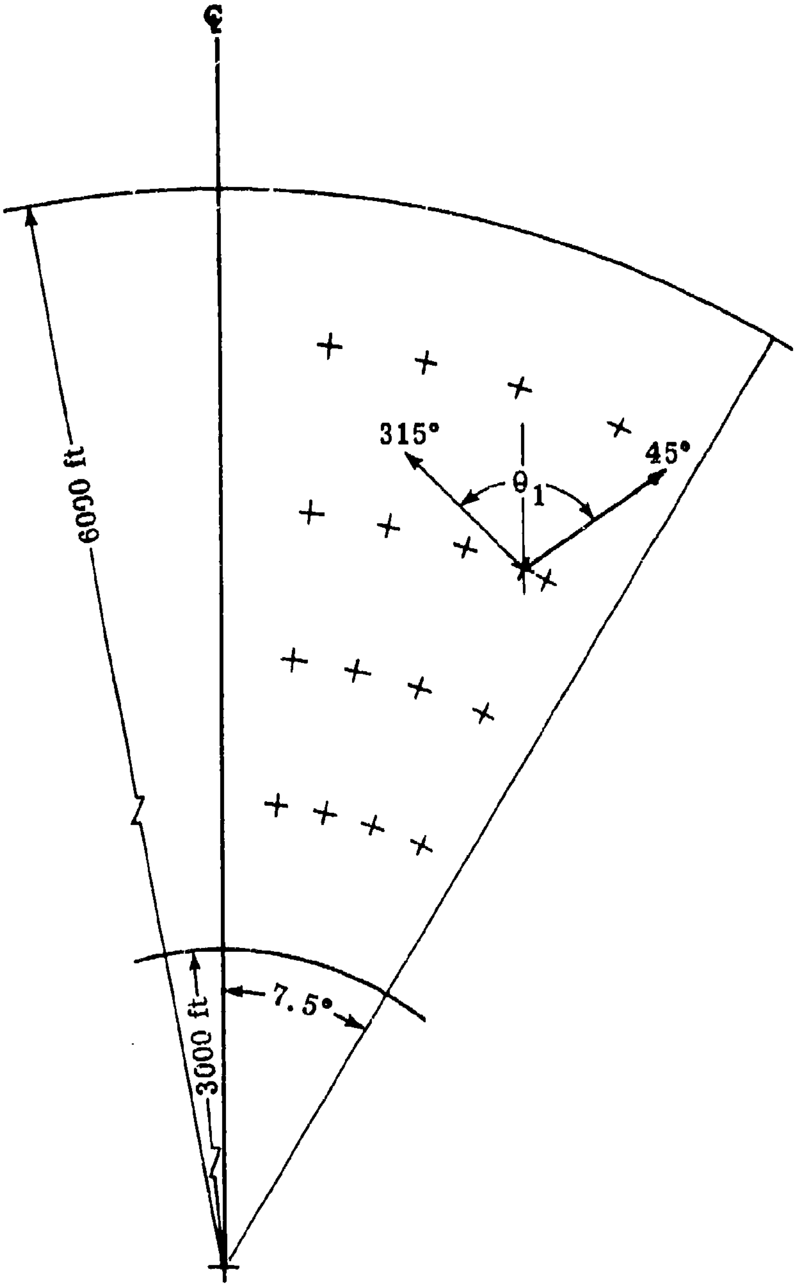

Figure 13. Region 22-θ_1, is defined for any (ρ, ω, θ) such that $3000\ \text{ft} < \rho \leq 6000\ \text{ft}$, $0° < \omega \leq 7.5°$ and $315° < \theta \leq 45°$